Chapter 3

Bottlenecks, genetic drift and the detection of selected variation

Deborah M. Leigh, Heidi Lischer, Frédéric Guillaume,
Christine Grossen, Torsten Günther,
Andreas Wagner, Lukas F. Keller

Abstract

The understanding of the ongoing evolutionary processes after a severe population decline, or bottleneck, is limited for wild populations. Contemporary knowledge is largely derived from population genetic theory or laboratory experiments, as the confounding effects of drift-driven false positive signals of selection have often hampered progression from theoretical to empirical investigations. In this study it was investigated if signals of selection could be accurately distinguished from genetic drift across 23 bottlenecked and reintroduced populations of Swiss Alpine ibex (*Capra ibex*). A set of neutral loci and loci under selection of varying effect size that contributed additively to a single trait were simulated. These loci were then also subjected to the reintroduction history of Alpine ibex populations, and were used to quantify the false and true positive rates of three different selection detection methods. High detection-accuracy, despite recent bottlenecks, was achievable for loci of large effect by combining loci identified by multiple methods and using a selection detection approach that incorporated environmental data. By applying these methods to Alpine ibex RADseq SNP data, a weak candidate selective response in relation to the depth of winter snow, an environmental variable known to drive high mortality events in the species, was identified.

Introduction

Identification of recent allele frequency change caused by selection is of great interest to evolutionary biologists. Understanding recent selective change can provide insights into adaptation and the propensity for responses to evolutionary pressures on different time scales (Whitlock and Lotterhos, 2015a). For conservation studies, insights into ongoing selective processes can help monitor a population's risk of becoming maladapted, as well as its long-term extinction risk, if a population or species is no longer able to respond to selection (Frankahm *et al.,* 2010).

In recent years, the ease of obtaining genome-wide SNP data due to high throughput sequencing has increased the number of studies looking for signals of selection at the genomic level (e.g. *Gasterosteus aculeatus*, Hohenlohe *et al.,* 2010; *Peromyscus maniculatus*, Linnen *et al.,* 2013; *Sarcophilus harrisii,* Epstein *et al.,* 2016). This interest in identifying selection from genomic data has caused a resurgence of attention to selection detection approaches such as *Fst* outlier analyses. *Fst* outlier analyses detect recent positive selective responses among populations of the same species, by identifying large differences in allele frequencies that result in unusually high values of *Fst* (Lewontin and Krakauer, 1973). High values of *Fst* are driven either by positive selection, i.e. selection for an allele that increases fitness, or by linkage to a positively selected region (Lewontin and Krakauer, 1973; Fay and Wu, 2000). Recent interest in outlier analyses has led to considerable analytical extensions of this approach and the development of software packages that are capable of identifying putatively selected loci (Beaumont and Nicolas, 1996; Vitalis *et al.,* 2003; Beaumont and Balding 2004; Foll and Gaggiotti, 2008; Excoffier *et al.,* 2009; Bonhomme *et al.,* 2010; Frichot *et al.,* 2013; Günther and Coop, 2013; Duforet-Frebourg *et al.,* 2014; Gautier, 2015a; Whitlock and Lotterhos, 2015a). Initially methods assumed an island-model demography (e.g. FDIST2, Beaumont and Nichols, 1996), where a species has several discrete populations that exchange migrants to the same extent, and therefore all populations are equally related (Wright, 1949; Latter, 1973). While this assumption is suitable for some species (e.g. the Galapagos Mockingbirds, *Mimus spp.,* Hoeck *et al.,* 2010), many species do not exhibit an island-model demography and therefore, have more complex relationships among populations. For example, it is quite common for geographically nearer populations to be genetically more closely related than distant populations due to a higher

probability of migrant exchange, leading to isolation by distance patterns of diversity (Wright, 1949). In an outlier analysis, violations of demographic assumptions can lead to a high number of neutral loci that are falsely identified as being under selection. This is because of unaccounted variance in the distribution of *Fst* due to shared history and relatedness of populations (Robertson, 1975a; Robertson, 1975b; Excoffier *et al.,* 2009). Therefore, outlier detection methods have since been extended to incorporate unequal relatedness across populations and can now be applied to a range of demographic scenarios (Bonhomme *et al.,* 2010; Günther and Coop, 2013).

Many early examples of local adaptation identified signals of selection using clines in phenotypes and the environment. For example, the cline of *Drosophila melanogaster's* ADH locus with latitude and the rapid reestablishment of this cline after the introduction of the species to new environments, is a classic example of an adaptive response to selection (Berry and Kreitman, 1993; Umina *et al.,* 2005). Such associations can arise when alleles, or the haplotype that they may be linked to, are selectively advantageous under certain environmental conditions (De Mita *et al.,* 2013). Selection detection analyses have been extended to enable the identification of local selective responses to such environmental clines (Coop *et al.,* 2010; De Mita *et al.,* 2013). Genetic-environment association analyses (GEA) pinpoint alleles that display repeated associations with an environmental variable (Lotterhos and Whitlock, 2015; Hoban *et al.,* 2016). Extensions to this approach have also been developed that can account for spurious associations, which may arise due to correlations in allele frequencies across closely related populations (Coop *et al.,* 2010). As a result, it is now possible to detect selection across a range of demographic scenarios through both outlier values of *Fst* and correlations of allele frequencies with environmental variables.

Despite the improvements made, the degree to which currently available GEA and *Fst* outlier methods can successfully accommodate non-island model demographies is still being tested. Recent simulation studies have shown that different selection detection methods can display large differences in the ability to (i) correctly identify loci under selection (true positives), (ii) falsely identify neutral loci as being under selection (false positives), and (iii) falsely identify selected loci as neutral (false negatives) (Bonhomme *et al.,* 2010; De Mita *et al.,* 2013; Lotterhos and Whitlock, 2014; Lotterhos and Whitlock, 2015). For example in populations that have undergone expansion from a glacial refugium, detection methods like those

implemented in the programs Bayenv 2.0 (Günther and Coop, 2013) or OutFLANK (Whitlock and Lotterhos, 2015a), display a low false positive rate and correctly identified several true positive loci, while the methods implemented in programs such as FDIST2 (Beaumont and Nicolas, 1996) or BAYESCAN (Foll and Gaggiotti, 2008) perform poorly (Lotterhos and Whitlock, 2014; Whitlock and Lotterhos, 2015a). This information is essential to ensure selection detection accuracy, yet many demographic scenarios are still unexamined. For example it is still largely unknown how recent population bottlenecks, periods of reduced population size, affect outlier analysis.

Recent population bottlenecks pose a challenge for the detection of loci under selection, because they are associated with strong genetic drift (Frankham *et al.,* 2010). The random allele frequency changes caused by genetic drift can lead to large allele-frequency differences between populations (Kimura 1955a; Kimura 1955b), creating outliers that may be mistaken as signals of selection (Lotterhos and Whitlock, 2014). Accordingly, strong genetic drift will likely increase the false positive rate when one tries to detect selection with genomic data (Klopfstein *et al.,* 2006; Nielsen *et al.,* 2007; Foll and Gaggiotti, 2008; Hofer *et al.,* 2009). Such false signals have previously hampered investigations of selection in humans (Sabeti *et al.,* 2006). Furthermore, many bottlenecked species may be under conservation management and therefore may have undergone reintroductions (Seddon *et al.,* 2012). Such reintroductions often lead to unequal and high relatedness among populations, as well as serial bottlenecks, when different source populations are used or when a source is used repeatedly (e.g. Biebach and Keller, 2009), posing an additional challenge for selection detection.

Examination of selection detection analysis in bottlenecked populations is currently limited to the effects of including a subset of populations that are bottlenecked (Foll and Gaggiotti, 2008). This was shown to cause a high false positive rate. It was therefore recommended to remove bottlenecked populations from such analysis (Foll and Gaggiotti, 2008). However, this is not always possible as there is often an interest in detecting loci under selection in bottlenecked populations. Species undergoing profound population declines are becoming increasingly common (Butchart *et al.,* 2010; Dirzo *et al.,* 2014). Well-studied genetic systems such as *Drosophila melanogaster* and humans are also known to have a history of population bottlenecks (Bamshead and Wooding, 2003; Haddrill *et al.,* 2005; Thornton and Andolfatto, 2006). Furthermore, outlier analyses have already begun to be applied to

bottlenecked populations (Pilot *et al.,* 2014; Funk *et al.,* 2016; Brüniche-Olsen *et al*., 2016). It is therefore important to have a detailed understanding of how population bottlenecks will affect signals of selection and the ability to identify selected loci.

In this study, selection identification in bottlenecked and reintroduced populations was explored. Recent studies have examined the performance of selection detection methods in terms of their true and false positive rates across a variety of simulated population demographic histories (e.g. De Mita *et al.,* 2013; Lotterhos and Whitlock, 2014; Lotterhos and Whitlock, 2015). Here, this type of evaluation was extended to bottlenecked populations. In addition, the effects of combining multiple selection detection methods were examined. Combining several methods may reduce the false positive rate but may also suffer from lower statistical power. To address these questions, genetic drift and selection were simulated using individual-based forward simulations that followed the known demographic history of reintroduced and bottlenecked populations of Alpine ibex (*Capra ibex*). The Alpine ibex were chosen as the basis for the simulations, because they are a recently bottlenecked species with excellent knowledge of their demographic history (Biebach and Keller, 2009). This knowledge allows for detailed simulations to generate SNP data containing realistic levels of drift and shared evolutionary history across populations. Furthermore, basing simulations on a species allows the application of our selection detection methods to an empirical data set. Within this study simulated data included both neutral and selected loci, enabling quantification of the rates of false positives, true positives and false negatives for different selection detection methods. This enabled the most successful selection detection methods identified by the simulations to be applied to an empirical RADseq SNP data set from Alpine ibex.

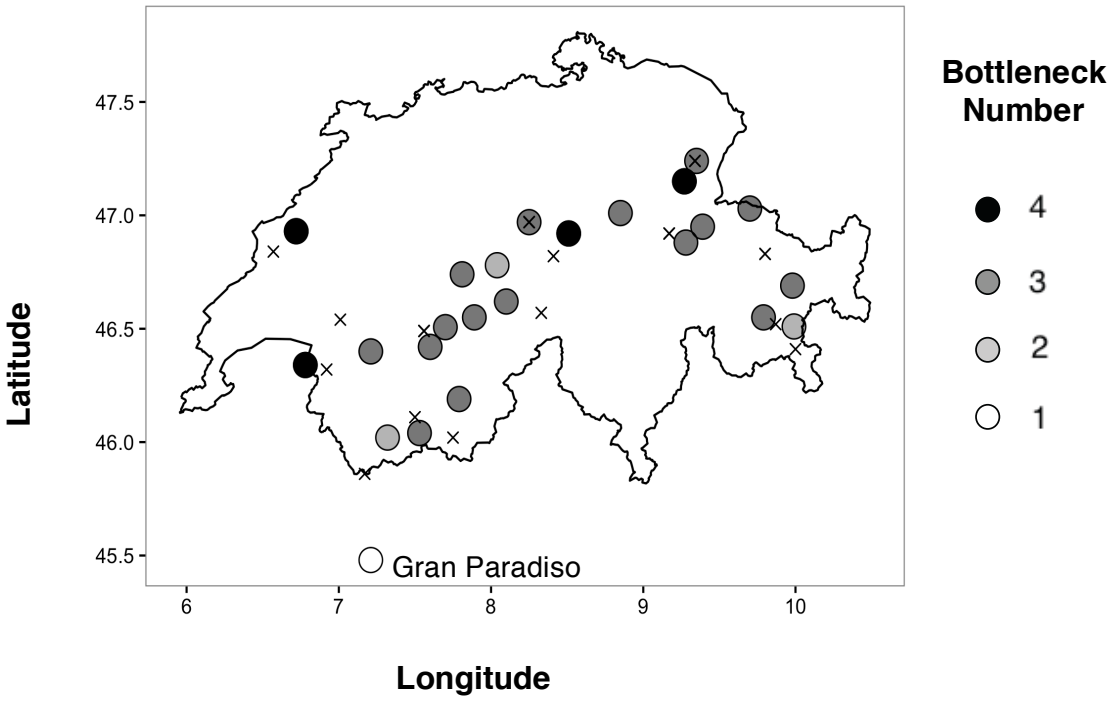

Figure 1: *A simplified depiction of the distribution of the 23 Alpine ibex focal populations, and the population-specific reintroduction history in Switzerland. All Swiss populations ultimately go back to a population in Gran Paradiso national park in Northern Italy (open circle), which is included in the figure but was excluded from the selection detection analysis. Reintroductions in Switzerland often used founder individuals from previously established reintroduced populations. As a result, many populations have experienced several serial bottlenecks. Within this figure, each circle represents a Swiss Alpine ibex focal population and the circle's shading indicates the number of bottlenecks each population experienced. The populations with darker shades indicate more bottlenecks occurred before population founding. Marked by a cross are the weather stations used to estimate the local environment experienced by each population. Coordinates were taken from BAFU (2016). The figure is simplified and only shows the core of each populations range. It excludes an additional bottleneck Swiss populations have experienced when Alpine ibex were bred in captivity to create a founder source for early reintroductions. All individuals from the population shown (excluding the Gran Paradiso) were sequenced with restriction site associated DNA sequencing in this study.*

Materials and methods

Alpine ibex demographic history

Alpine ibex underwent a prolonged decline starting in the 16th century due to overhunting. Only a single population of an estimated 100 or fewer individuals survived this crash in the Gran Paradiso region in Northern Italy. Royal protection in the 19th century enabled the population to grow to 3000-5000 individuals. Reintroductions of Alpine ibex from the Gran Paradiso region into Switzerland began in 1906. Detailed demographic records were kept as part of the reintroduction program in Switzerland. Information that was recorded included the origin of founder individuals (often coming from previously reintroduced populations), the number and gender of founders, and the year an individual was moved. In addition, annual records of the counts of the number of animals alive in spring were collected for many reintroduced populations (Stuwe and Grodinsky, 1987; Stuwe and Neivergelt, 1991; Biebach and Keller, 2009). This reintroduction program was very successful, and to date more than seventeen thousand Alpine ibex are present in the Swiss Alps (Shackleton and Group ISCI, 1997; BAFU, 2015). The focal populations used in this study are shown in Figure 1. All figures were produced in ggPlot2 and ggMap (Wickham, 2009; Kahle and Wickham, 2013) using R version 3.3.0 (R, 2016).

Simulating the Alpine ibex history

To generate simulated SNP data for the selection detection analysis, forward time simulations were conducted using Nemo (version 2.3.44; Guillaume and Rougemont, 2006). Details of the simulations can be found in the supplementary material (S1). Briefly, 15 replicate simulations of the Alpine ibex reintroduction history were conducted. In each simulation all 23 populations sampled for restriction site associated DNA sequencing (RADseq) were simulated. In order to accurately simulate these populations, an additional three populations that were founder sources for the focal populations were also simulated. Therefore, 26 populations were simulated in total. For each individual 60 thousand neutral loci and an additional 30 loci under selection that contributed to one artificial quantitative trait were simulated. 4000-5000 of the neutral loci remained polymorphic after the bottlenecks. All loci were spread across 30 chromosomes. The 30 loci under selection were divided equally into three categories, each representing a specific positive or negative value

(allelic value) that the locus contributed to an individual's trait value. Ten loci with large contributions to each trait (allelic value ±0.1), 10 of moderate (±0.01) and 10 of minor contribution (±0.001) were simulated. Therefore, maximum trait values of ±2.22 was achievable if all loci were homozygous for alleles of only negative or positive affect. Neutral loci and loci under selection were allowed to reach mutation-selection-drift equilibrium during a "burn-in" of 50 thousand generations in a single population that represented the Gran Paradiso populations, after this time a bottleneck was applied to the population. The trait optimum value during the burn-in was held at zero to maintain alleles of both negative and positive affect. To create a shift in allele frequencies and an association of frequencies with the environment in reintroduced populations, the trait optimum in reintroduced populations was varied to either zero, -2 or +2. This value is hereafter referred to as the 'true optimum.' Simulated optima in the reintroduced populations were chosen to reflect the natural variation in environmental conditions across the Alpine ibex populations, and were therefore based on the snow conditions each population experiences. Snow conditions were chosen for this analysis, because they have previously been shown to affect Alpine ibex population dynamics and vary dramatically across sites (detailed in S1 and S2) (Jacobsen *et al.,* 2004; Grøtan *et al.,* 2008). Based on the relative winter snow depth of each population in comparison to the Gran Paradiso populations, which had a simulated optimum of zero, nine reintroduced populations were given an optimum of zero, 14 reintroduced populations an optimum of +2, and two reintroduced populations an optimum of -2. This shift from zero to ±2 created positive selection towards the true optimum in a subset of the reintroduced population at varying strengths across the 30 loci. To calculate selection coefficients (*s*) at these loci, a random idealised population was simulated in R (R Core Team 2016). The simulated population experienced an optimum shift of zero to two, and a phenotypic variance (V_p) of 0.07, as well as a selection variance estimate (ω^2) of 1.5 (as in the Nemo simulations) were assumed. The average *s* across all genotypes in this population was calculated for each allelic value. The shift in optimum translated into selection coefficients of 0.13 (allelic value ±0.1), 0.013 (allelic value ±0.01), and 0.0013 (allelic value ±0.001). Simulations in which any population had gone extinct (n=4) were discarded from further analysis resulting in 11 simulated data sets. The simulated genotypes of the final generation were used for selection detection, and only polymorphic SNPs were included in the simulated data from this time point. To

mimic the available RADseq data (described below), 10 individuals were randomly chosen from each of the 23 populations that were sequenced with RADseq, 20% of genotypes were randomly set to "missing" due to missing data in the RADseq genotypes and singletons were removed (vcftools; Danecek *et al.,* 2011). PGDspider (version: 2.0.9.2) and custom scripts were used to convert Nemo output into input for the selection analyses (Lischer and Excoffier, 2012).

Screens for signals of positive selection

To identify candidate SNPs under selection, selection detection analyses were conducted for both the simulated data sets and the empirical Alpine ibex RADseq data. To detect signatures of selection Bayenv 2.0 (Günther and Coop, 2013), Baypass 2.1 (Gautier, 2015a), and OutFLANK (Whitlock and Lotterhos, 2015a) were used. These three programs were chosen as they have previously been shown to perform well in species with complex patterns of relatedness across populations that arise from demographies which do not follow a simple the island model (Günther and Coop, 2013; Lotterhos and Whitlock, 2014; Gautier, 2015a; Whitlock and Lotterhos, 2015a). Outlier analyses detect selection by identifying loci with larger differences in allele frequencies between populations than expected based on the population average. Bayenv 2.0 and Baypass2.1 utilize a modified *Fst*-like statistic called X^TX that is corrected for shared population history (Günther and Coop, 2013; Gautier, 2015a). Outflank utilizes an *Fst* statistic called *F'st*, a metric based on Wright's *Fst* statistic without corrections for a finite sample size (Whitlock and Lotterhos, 2015a). These three methods are referred to as *Fst*-like approaches throughout this study. Bayenv 2.0 and Baypass2.1 also detect selection through identifying repeated associations in allele frequencies with an environmental variable these methods are hereafter referred to as GEA (genetic-environment association) approaches (as in Hoban *et al.,* 2016). To detect loci that may be subject to selection, loci identified as potentially under selection by each program individually and across multiple programs were examined. Loci identified across two (double positives) or all three programs (triple positives) (Bayenv 2.0, Baypass2.1, and OutFLANK) as *Fst* outliers were of interest for the *Fst*-like analyses. OutFLANK is unable to detect selection using the GEA approach, hence for the consensus analysis (double or triple positives) of the environmentally correlated loci identified by the GEA approach from Bayenv 2.0 and Baypass2.1 were used in addition to the *Fst*-like outliers from OutFLANK.

When examining loci flagged as putatively under selection, a true positive was considered to be a simulated locus under selection that was correctly identified as being under selection. A false positive was considered to be a simulated neutral locus that was wrongly identified as being under selection. The percentage of all loci identified as under selection that were true positives (the true discovery rate), was used as a metric of the accuracy and reliability of selection detection. To place the results in the context of other simulation studies, the true and false positive rates and the false discovery rate were also calculated (Lotterhos and Whitlock, 2014). The true positive rate is the number of true positives divided by the total number of simulated loci under selection. The false positive rate is the number of false positives divided by the total number of neutral loci. The false discovery rate is the proportion of all loci identified as being under selection that were false positive loci (Lotterhos and Whitlock, 2014). False negative loci, i.e. simulated loci under selection that were not identified as so, were also considered in this study. The false negative rate, the number loci that failed to be identified as under selection divided by the total number of simulated loci under selection, was calculated for each category of simulated selected loci (i.e. allelic values of 0.1; 0.01; 0.001). In summary, the true and false discovery rates are defined relative to the total number of loci identified as under selection. The false positive, false negative, and true positive rates are relative to the total number of loci of each SNP type in the data set as a whole. All values presented are averages of the selection detection analyses that were conducted independently across 11 simulated datasets.

All environmental data used in the GEA analyses in Bayenv 2.0 and Baypass2.1 were obtained from MeteoSwiss (2016). For each population, data from the closest meteorological station available (Figure 1, Section S1 and S2) were used to obtain estimates of the average weather condition experienced since a given population was founded, or since records began, if weather data collection began after population founding. The environmental variables in the analyses were divided across winter and summer and included air temperature (°C), daily precipitation (mm), and snow depth measures (cm). Further details are available in the supplementary material (section S1 and S2). Since the simulations were intended to mimic real Alpine ibex populations, the corresponding weather data were included as environmental covariates in the Bayenv 2.0 and Baypass2.1 analyses of the simulated data. In

addition, each simulated population's true optimum was also included as an environmental covariate in the analysis of the simulated data (Table S1).

For both the simulated and the Alpine ibex RADseq data, Bayenv 2.0 was executed three times with 2 x10^5 Markov-Chain-Monte-Carlo (MCMC) iterations both for the estimation of covariance in allele frequencies between populations, as well as for the tests for outliers and environmental correlations in allele frequencies (Blair *et al*., 2014). All polymorphic SNPs were used for the estimation of the covariance matrix of allele frequencies. This covariance matrix estimates relatedness among all populations and is used to correct for shared population history when identifying loci under selection (Günther and Coop, 2013). Loci simulated to be under selection were not excluded for the estimate of the covariance matrix in the analysis of the simulated data (De Mita *et al.,* 2013), because their counterparts in the empirical data set, i.e. SNPs under selection in the Alpine ibex, cannot be excluded from empirical data sets (as in Lotterhos and Whitlock, 2014). The final covariance matrix estimated for each of the three runs of 2 x10^5 MCMC iterations in Bayenv 2.0 was used in the selection detection analysis. Putatively selected SNPs were determined in two ways: 1) for the GEA approach SNPs were considered to be under selection if their Bayes factor (BF) was above three for all replicate runs, which indicates substantial support for a SNP being under selection (e.g. Nadeau *et al.,* 2016). For all loci, Bayenv 2.0 also calculates the Spearman's rho correlation coefficient between allele frequencies and each environmental variable. For loci considered to be subject to selection, loci should also have high values of this correlation coefficient (Günther and Coop, 2013). For this reason, GEA loci that exceeded the BF threshold had to have an average Spearman's rho value in the top and bottom 2.5% of all SNPs across the three runs. This ensured that the trend could not be driven by one extreme population (Günther and Coop, 2013). 2) A different criterion for outliers was for the *Fst*-like approach. Here, outliers included SNPs with an X^TX value among the top 100 ranking SNPs across all three runs (Günther and Coop, 2013). X^TX is a differentiation statistic similar to *Fst* and high X^TX values signal putative selection. X^TX values were corrected for the shared population history estimated by the covariance matrix across populations, to ensure that shared population history does not create artificially high values of *Fst* (Günther and Coop, 2013).

Baypass2.1 was run under default conditions (20 pilot runs of 1000 MCMC iterations, 5000 MCMC iterations for "burn-in") using the auxiliary model and an estimate of the covariance matrix produced by the core model (Gautier, 2015a). The covariance matrix was also used to correct for shared population history in the analysis to detect loci under selection. The thresholds used to identify loci putatively under selection were taken from the best-practice tutorial accompanying Baypass2.1 (Gautier, 2015b) and included two criteria: 1) For the GEA approach SNPs were considered to be under selection when they showed a 10 x $\log_{10}$ Bayes factor (db) greater than 20 (Gautier, 2015a). 2) For the *Fst*-like approach X^TX outliers were determined by creating three Baypass-simulated data sets of 1000 SNPs to determine a 99% threshold for X^TX values and SNPs in the top 1% of X^TX values were considered outliers (Gautier, 2015b). Allele frequency distributions of SNPs used in the Baypass-simulated data were drawn from the input data set. To ensure that no singletons were present in the Baypass-simulated datasets, a minimum minor allele frequency of 0.005 was applied. Furthermore in the Baypass-simulated datasets, correlations of SNPs with the environmental variables were set to zero, to ensure all SNPs behaved neutrally (Gautier, 2015b).

In OutFLANK, outlier SNPs were first identified using the default settings (Whitlock and Lotterhos, 2015b). Under these settings, OutFLANK assesses if a locus is under selection by creating a χ^2 distribution based on the distribution of *F'st* values. This χ^2 distribution is used as a null distribution to calculate p-values for each locus's likelihood of being an outlier. Q-values are then based on the right tailed p-values (q-values are similar to p-values but are corrected for the false discovery rate). A SNP is defined as an outlier if it has a q-value of less than 0.05 (Storey and Tibshirani, 2003; Whitlock and Lotterhos, 2015a). A SNP had to show a heterozygosity of greater than 10% to confirm its outlier status (Whitlock and Lotterhos, 2015b). Due to a very high false negative rate that arose when using the default settings, a less conservative approach was also applied, defining an outlier as a SNP with a right tailed p-value of less than 0.05. This approach can be considered uncorrected for the false discovery rate (Storey and Tibshirani, 2003).

False positives and linkage

The initial criteria to identify true positive loci in the simulated data sets only considered loci directly under selection. However, because all simulated loci were

placed on a chromosome it is possible that neutral loci may show signals of selection due to linkage with selected loci. Such loci can be considered either false positives, as they themselves are not directly under selection, or a marker that correctly signals an allele frequency change driven by selection due to linkage to a locus under selection. To estimate the number of such loci, linkage disequilibrium was calculated between all loci identified as under selection in the simulated data and those loci actually simulated as under selection. Vcftools was employed to calculated the r^2 values (Danecek *et al.,* 2011). Linkage disequilibrium rather than information on physical linkage was used in this analysis because location information was unavailable for the simulated markers (this was due to a large-scale computational problems that lead to loss of data, physical linkage information is being regenerated for publication). It is therefore assumed that linkage and linkage disequilibrium are correlated in this analysis, though it is possible that high linkage disequilibrium exists between physically distant SNPs, and that physically close SNPs may not be in linkage disequilibrium (Ardlie *et al.,* 2002). In this analysis a threshold of $r^2 = 0.25$ was used to indicate high linkage disequilibrium. Therefore all neutral loci that were identified as being under selection and that exhibited an $r^2 > 0.25$ with a simulated locus that was under selection were summed for this analysis.

Simulations and perfect knowledge of the environment:

For the simulated data set perfect information about the selective optimum to which loci that were under selection loci were evolving towards, was included in the GEA analysis (section S1, Table S1). This analytical situation is unlikely to arise in nature. Consequently, for the Bayenv 2.0 and Baypass2.1 GEA analyses the effect of removing this information on the number of true positives identified was assessed. To this end, all loci whose allele frequencies were only significantly associated with the true environmental optimum were removed and the reduction in the true positive and discovery rate was calculated. Each locus can display multiple significant correlations in allele frequencies with environmental variables, this means that only loci with one significant environmental association (with the true environmental optimum) were removed. The average reduction in the number of true positives identified was then calculated.

RAD sequencing:

To apply the selection detection methods to an empirical data set, genomic DNA of 304 Alpine ibex from 23 reintroduced populations was sequenced with RADseq (Figure 1) (Baird *et al.,* 2008; Etter *et al.,* 2011). After SNP filtering (described in section S3), individuals with more than 50% SNPs missing were removed. This resulted in a sample of 210 individuals, of which 5 were additionally removed because they were sampled in duplicate or displayed biologically impossible heterozygosity. For selection detection all singletons, variants on the X-chromosome, and variants that could not be mapped to the annotated goat genome (CHIR_1.0, Dong *et al.,* 2013) with MUMmer (Delcher *et al.,* 2002) were removed, which resulted in a final data set of 6864 SNPs. SnpEFF (Cingolani *et al.,* 2012) was used to annotate the putative SNP effect. A database for this analysis was created using the goat genome (version CHI1)(*Capra hircus*, Dong *et al.,* 2013).

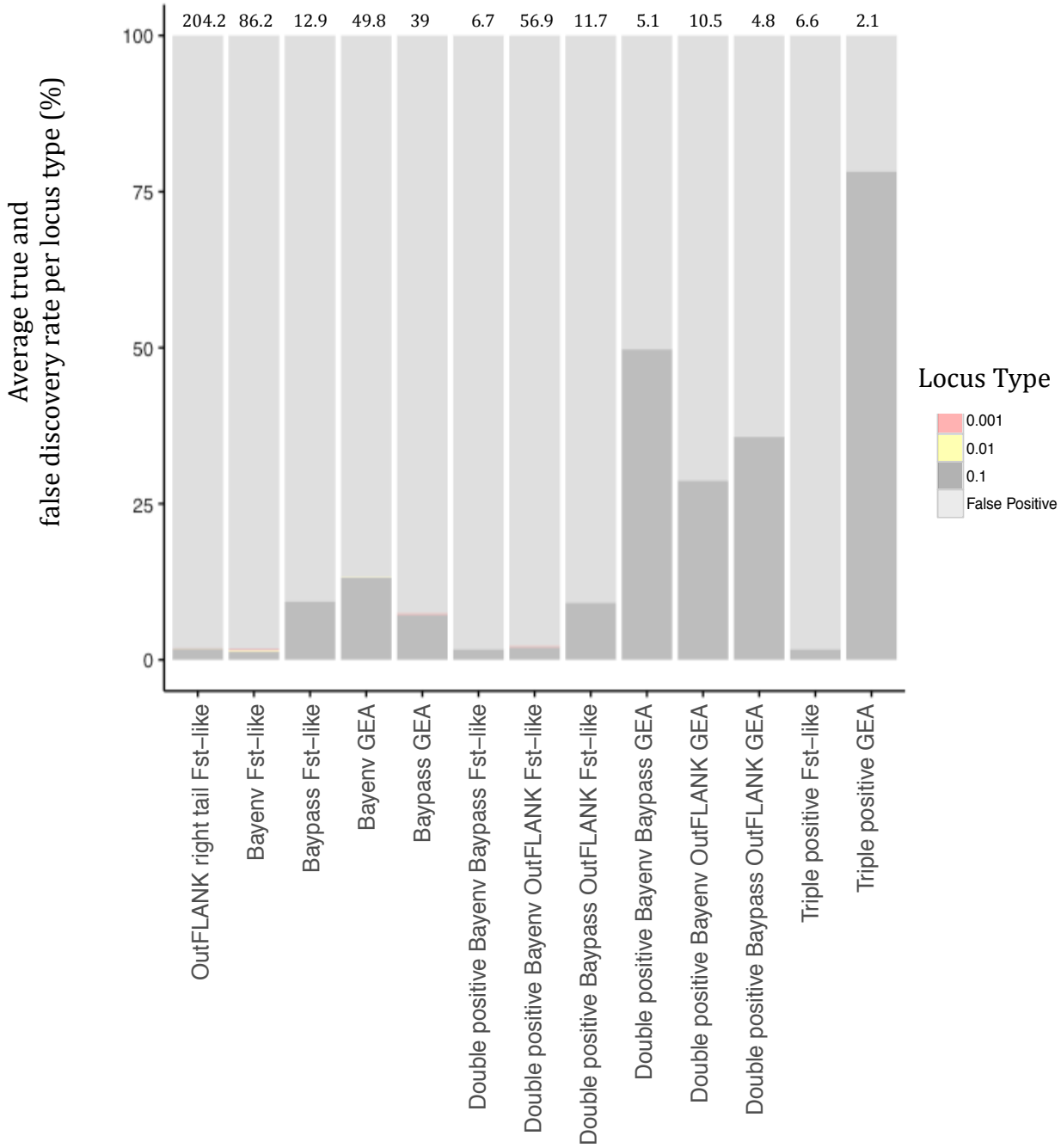

Figure 2: *Simulated loci identified as under selection by each detection approach and by the consensus of multiple methods (double and triple positives). The values shown are the true and false discovery rates. True positives are loci simulated under selection that were correctly identified as being under selection. False positives are neutral loci that were identified as being under selection. The numbers at the top of each bar represent the average number of all loci detected by each method. Loci are divided into those that are true positives for each simulated allelic value (red, yellow or dark grey) or are false positives (light grey). For the true and false discovery rates, the number of each type of locus is divided by the number of all loci detected as being under selection. Loci with allelic values of 0.01 or 0.001 all have true discovery rates between zero and 0.32%. GEA approaches identify selection by looking for an association between an environmental variable and allele frequencies. Fst-like approaches use high differentiation in allele frequencies between populations to identify selection. The triple positive approach overlapped the outliers identified by Bayenv 2.0, Baypass2.1 and OutFLANK. All values shown are averages across 11 simulated data sets (standard deviations of results from each detection method in Table S3).*

Results

Screens for signals of positive selection

In this study, 4000 to 5000 polymorphic neutral SNPs and 30 SNPs subject to selection were simulated that followed the Alpine ibex demography. Bayenv 2.0, Baypass2.1, and OutFLANK were then used to identify selection in this simulated dataset using *Fst*-like statistics (X^TX and *F'st*) and genetic-environment association (GEA) approaches. OutFLANK, using the default settings, identified no loci subject to selection. Therefore, OutFLANK outliers were identified using less conservative settings, considering outliers detected by the right tailed p-value value (see methods) for further analyses. The simulated data is discussed in the order of true discovery rate, true positive rate, false positive rate, false discovery rate and false negatives rate (defined in methods).

Analyses of simulated data revealed that GEA approaches showed the highest true discovery rate (Figure 2, Table S3), especially for the double and triple positive methods (defined in methods). For the double positive GEA method, the true discovery rate was between 28.7% and 49.7% and up to 78.2% for the triple positives. The true discovery rate for the *Fst*-like approaches was lower ranging from 1.7 to 9.1% for the double positive method (Figure 2). For the triple positive *Fst*-like method only 1.7% of the outlier loci were true positives. All selection detection methods struggled to detect true positive loci that were under selection with an allelic value of 0.01 or below. These loci always had a negligible true discovery rate of between 0 and 0.32% (Figure 2) and a true positive rate of 0-2.7% (Table S4). For loci with an allelic value of 0.1, the true positive rate for each method ranged from 1.8 to 60%. The false positive rate for each method ranged from 0.01% for the GEA triple positive method to 4.13% for the OutFLANK right tailed p-value approach. For each program individually, the false discovery rate ranged from 86.8% for the GEA approach in Bayenv 2.0 to 98.2% for *Fst*-like approach in Bayenv 2.0 (Figure 2). For the double positive methods the false discovery rate ranged from 50.3% (for the Bayenv and Baypass GEA approach) to 98.3% (for the Bayenv Baypass *Fst*-like approach). The triple positive methods had the greatest disparity in the false discovery rate, with the triple positive GEA method exhibiting a false discovery rate of 21.8% relative to 98.3% for the triple positive *Fst*-like method.

In the simulations, the allelic values and hence the strength of selection experienced by each selected locus were not equal. The ten loci of the largest effect (allelic value 0.1) were under much stronger selection (s=0.13) relative to those with allelic values of 0.01 (s=0.013) or 0.001 (s=0.0013) and therefore were more likely to have undergone large enough allele frequency changes to be identified as being under selection. The false negative rate for loci with an allelic value of 0.1 ranged from 40% to 98.2% (Figure 3,Table S4). For loci of weaker effect almost no loci were identified as outliers and the false negative rate ranged from 97.3% to 100%.

False positives and linkage

Because the simulated loci were dispersed over 30 chromosomes, false positive neutral loci may arise due to linkage with a simulated locus under selection. The linkage disequilibrium (r^2) was calculated between all neutral loci and loci simulated under selection. The average degree of linkage disequilibrium across all loci was very low and ranges only from 0.0094 to 0.0118 across simulations. However, higher levels of linkage disequilibrium were observed among some pairs of false positive and loci under selection. However, no more than 0% to 3.9% of all false positive loci identified by the different selection detection methods were in high linkage disequilibrium ($r^2 > 0.25$) with a locus simulated under selection (Table S6). The triple positive *Fst*-like method detected an average of 2.8% of false positive loci that were in high linkage disequilibrium with a selected locus and the triple positive GEA method identified no such loci.

Simulations and perfect knowledge of the environment

Within the simulated data set, a risk exists that the higher accuracy identified by the GEA approach is due to perfect knowledge of the true optimum. After removing loci that were only identified because of associations with the true optimum, the Bayenv 2.0 GEA approach lost on average 0.9 true positive loci and the Baypass2.1 approach lost no such locus. The decrease in the number of simulated loci that were correctly under selection identified by Bayenv 2.0 corresponded to a 14% decrease in the average number of true positives identified. The relatively small loss of true positive loci was likely caused by detection of the association through other, correlated, environmental variables. Because loci can be significantly associated with multiple

environmental variables, it is possible that environmental variables correlated with the true optimum offer sufficient power to detect most loci subject to selection.

Alpine ibex RADseq data and signals of selection

The Alpine ibex RADseq dataset comprised 204 individuals (between 6 to 11 per population) and 6864 polymorphic SNPs after quality filtering (methods and S1). The triple positive GEA method that detected the highest percentage of true positive loci in the simulations identified no loci under selection in the RADseq dataset. The methods involving double positive GEA approach identified fourteen putatively selected loci across all three double positive methods (Table S7). Allele frequencies across the loci were associated with snow conditions and temperature. Six loci showed multiple associations, though this was often with the same environmental variable that had been divided into seasonal averages. The triple positive *Fst*-like method, which had a lower true positive rate in the simulations than the GEA approaches, identified 5 loci as putative outliers (Table S8). Overall, loci under selection could not be identified confidently in the empirical RADseq data set.

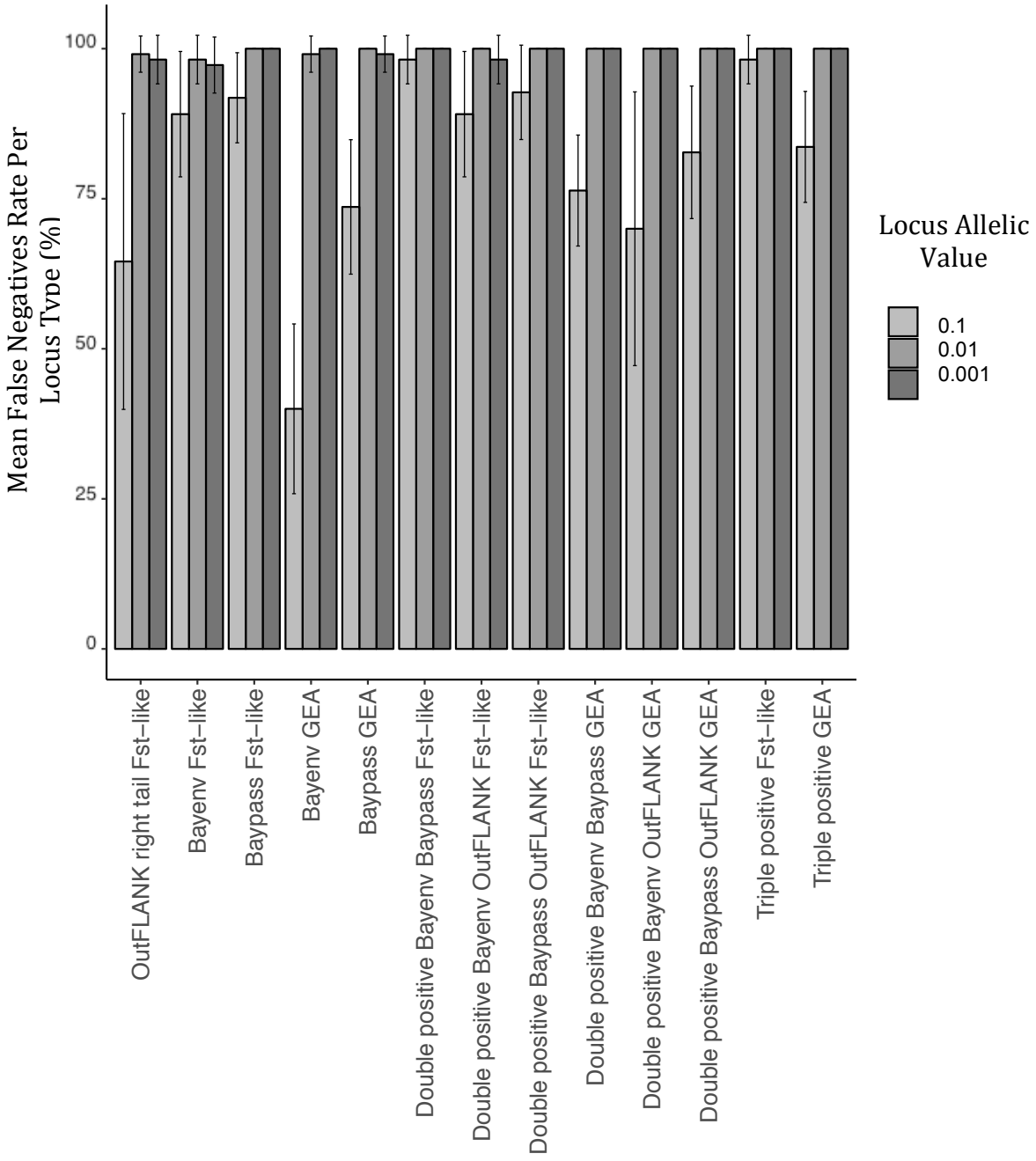

Figure 3: *The false negative rate of each selection detection method. 30 loci were under selection in the simulations, divided into three sets of 10 for each allelic value. Different grey shades correspond to different allelic values i.e. the value a locus can contribute to a trait. Values correspond to averages across 11 simulations (± one standard deviation).*

Discussion

In this study, the accuracy and feasibility of detecting recent signals of positive selection was assessed for a species that has undergone a global bottleneck, as well as repeated population bottlenecks as a consequence of a reintroduction program. Neutral loci and loci under selection were simulated and followed the well-documented population history of 23 bottlenecked and reintroduced populations of Alpine ibex. To create selective responses in the simulated populations, at the beginning of the reintroductions an environmental optimum (0 or ±2, see methods) was set for each population. This optimum reflected the natural variation observed in winter snow conditions in Alpine ibex populations. Thirty loci under selection that contributed to a single trait adapting towards this optimum were simulated along side 4000-5000 neutral loci. The true positive, false positive, false discovery, true discovery and false negative rates of three selection detection methods were quantified under this demography. The true and false discovery rates are metrics relative to the total number of loci identified as potentially under selection. The false positive, false negative, and true positive rates are defined relative to the total number of loci of each marker type in the data set as a whole. True positives are loci simulated under selection correctly identified as being under selection and false positives are neutral loci incorrectly identified as being under selection. Signals of selection of any one locus were identified 1) based on correlations between allele frequencies and an environmental variable (GEA approaches), which can arise when an allele is favoured across an environmental gradient (Hoban *et al.,* 2016), and 2) based on strong differentiation of allele frequencies relative to the average between populations (*Fst*-like approaches), which can signal recent local responses to positive selection (Hoban *et al.,* 2016). The simulated data revealed a clear difference in the performance of the GEA detection approach compared to the *Fst*-like approaches. The GEA approaches had a generally higher accuracy (lower false positive rates) and correctly identified more simulated loci that were under selection across all methods (higher true positive rates and true discovery rates). These differences are discussed in comparison to other studies below.

Simulated data and selection detection accuracy:

A clear difference in the performance was found for the different selection detections methods in this study. The GEA approaches generally had a lower false positive rate (Table S5) and a lower false negative rate (Figure 3) than the *Fst*-like outlier approaches. Furthermore, the GEA approaches had consistently higher true discovery rates (Figure 2). The highest true discovery rate for the *Fst*-like approaches was only 9.3% (Baypass), relative to the highest value for GEA approaches of 78.2% (triple positive). The true positive rate for loci of the largest effect was also lower for the *Fst*-like approaches and ranged from 1.8 to 35.5%, compared to 16.4 to 60% for the GEA approaches. Previous studies comparing these approaches have not found a consistent pattern of relative performance: GEA approaches have been found to perform better or worse than *Fst*-like approaches (De Mita *et al.,* 2013; Lotterhos and Whitlock, 2015). These differences in performance appear to depend on demography. For example while the GEA approach within Bayenv2.0 performs well in expanding populations (Lotterhos and Whitlock, 2014), the power of Bayenv2.0's GEA approach is lower than that of *Fst*-like approaches in populations that display strong isolation by distance (Lotterhos and Whitlock, 2015). This is thought to be due to correction for shared population history, which causes a substantial loss of power due to correlations between demography and environmental variables in populations that display isolation by distance (Lotterhos and Whitlock, 2015). It is discussed below why unequal performance of the approaches may have arisen in the demography that was simulated, with bottlenecked and reintroduced populations.

Genetic drift can cause allele frequency differences that mirror signals of selection (Kimura 1955a; Kimura 1955b; Lotterhos and Whitlock, 2014). Because strong genetic drift is expected after a bottleneck, a relatively higher false positive rate should be observed in bottlenecked populations relative to the analysis in populations of constant size. It has previously been shown that a false positive rate up to 10% can be expected if bottlenecked populations are included as a subset of populations in a selection detection analysis (Foll and Gaggiotti, 2008). All the populations in the study were bottlenecked and a false positive rate above 10% was not observed. In fact, the highest false positive rate identified was for *Fst*-like approaches with OutFLANK's right tailed p-value approach (4.13%). The relatively high false positive rate for this approach was not unexpected, as the right tailed p-value approach removes the false discovery rate correction otherwise implemented in

OutFLANK. The right tailed p-value approach was used in OutFLANK because the recommended, and more stringent, q-value threshold identified no true positives in the simulated data. Although the false positive rates across all methods were at the lower end of what has previously been shown when including bottlenecked populations (Foll and Gaggiotti, 2008), they were still considerably higher than those previously identified for populations of constant size. For example, the false positive rate of the *Fst*-like approach in Bayenv2.0 was 1.76% in the simulations compared to less than 0.5% previously identified in populations of constant size (Lotterhos and Whitlock, 2014). Consequently, though the results suggest that a lower false positive rate for *Fst*-like approaches than previously shown is possible, for bottlenecked populations the rates will be higher than those estimated for populations of constant size. This is likely due to strong genetic drift generating false positive signals of selection.

On average, GEA approaches had lower false positive rates (ranging from 0.01 to 0.91%) than *Fst*-like approaches. The lower false positive rates of the GEA approach relative to the *Fst*-like approaches may be caused by a reduced frequency of drift-driven false positives under the Alpine ibex demography. False positives due to drift and population history are possible in GEA approaches for selection (Rellstab *et al.,* 2015; Hoban *et al.,* 2016). However, these arise from correlations between population history and the environment. For example, range expansion along an environmental gradient can create clines in allele frequencies that mimic selection but are solely due to drift (Keller *et al.,* 2009). Alternatively, false positives may arise when locally shared population history is confounded with locally similar environments, creating a false correlation in allele frequency with environmental values (Hoban *et al.,* 2016). Due to the human-assisted origin of reintroduced populations like the Alpine ibex, closely related populations can be geographically distant or in locations with different environmental averages (Biebach and Keller, 2009). Therefore, false environmental correlations are less likely to arise in the reintroduced populations that were simulated in this study, and this has likely lead to the lower false positive rate of GEA approaches relative to the *Fst*-like approaches that was observed.

All methods included in this study had a true positive rate below 60% for loci of the strongest effect (allelic values) and below 2.7% for loci of weaker effect. The false negative rates ranged from 40% to 98.2% for loci with an allelic value of 0.1 (Table S4), and from 97-100% for loci of weaker effect. The ten loci of strongest

effect experienced a selection pressure (selection coefficient) of approximately 0.13, and the loci of weaker effect were estimated be under a selection pressure of roughly 0.013-0.0013. The lower true positive rate for loci with smaller allelic values is expected, because the failure of selection detection methods to identify loci under weaker selection pressures has been identified repeatedly (e.g. Biswas and Akey, 2006; Kalsson and Moen, 2010; Narum and Hess, 2011; Kemper *et al*., 2014; Lotterhos and Whitlock, 2015). However, in the study loci of weaker effect were identified much less frequently than in previous studies. This was despite the use of comparatively high selection coefficients (s=0.0013-0.013). For example, in this study loci with an allelic value of 0.01 or 0.001 were not identified by the double or triple positive method (Figure 2). Previous studies have identified a minimum of 11% of loci under weak selection pressures (*s*=0.005) with two or more selection detection methods (Lotterhos and Whitlock, 2015). This difference may be due to the shorter simulation time span of the simulations. The simulated populations were tens of generations old, while previous studies have simulated populations for thousands of generations (Lotterhos and Whitlock, 2015). Therefore smaller allele frequency differences are expected for loci under weak selection pressure in the simulations, likely reducing the number of selected loci correctly identified.

The true positive rate for loci under strong selection pressure (s=0.13) was also generally below that previously shown for populations of constant size. Lotterhos and Whitlock (2015) compared the number of true positive loci identified by the *Fst*-like approaches of Bayenv 2.0 and PCAdapt (Duforet-Frebourg *et al.,* 2014) to the GEA approaches of Bayenv 2.0 and LFMM (Frichot *et al.,* 2013) under a range of demographic scenarios. The lowest true positive rate they observed for these comparisons was 43% for the *Fst*-like approaches and 58% for the GEA approaches. These rates occurred for populations with an island model demography (Lotterhos and Whitlock, 2015). De Mita *et al.,* (2013) found a true positive rate of ~100% across multiple demographic scenarios for the Bayenv (Coop *et al.,* 2010) GEA method. In this study, the true positive rate observed for loci of the largest effect was 1.8% to 35.5% for the *Fst*-like approaches and 16.4% to 60% for the GEA approaches. High rates of genetic drift are likely responsible for the lower true positive rates observed. In addition, the use of a polygenic trait may have contributed to the generally lower true positive rate observed relative to previous studies. Polygenic traits have previously been shown to be difficult to identify as smaller changes in allele

frequencies are expected (Kemper *et al.*, 2014; Berg and Coop, 2014). It should be noted that, low true positive rates translate into high false negative rates. Therefore, if selection is not identified in bottlenecked populations this should not be considered evidence for an absence of selection.

The selection pressures that were simulated were strong. The selection pressure experienced by loci of strong effect was 0.13. It was 0.013-0.0013 for loci of weaker effect. However, very weak selection pressures could not be simulated because reintroduced populations have a small effective population size and those below the reciprocal value of two times the effective population sizes, will not illicit a change in allele frequencies (Lynch, 1996; Frankham *et al.,* 2010). Selection pressures as high as were simulated may be infrequent in nature. As a result, it is likely that studies of bottlenecked populations may have true positive rates closer to or below, those identified for the weaker loci in the simulations. However, very strong selection pressures do occasionally occur in nature. Estimates of selection coefficients have been shown to reach 0.59 for selection against homozygotes of the beak-size determining HMGA2 allele in medium ground finches (*Geospiza fortis*) during a drought that selected against large beak sizes (Lamichhaney *et al.,* 2016). Similarly, estimates from rapid evolutionary change of stickleback (*Gasterosteus aculeatus*) morphology from a salt- to a freshwater adapted phenotype revealed selection coefficients of up to 0.52 after an earthquake in Alaska isolated populations from the ocean (Lescak *et al.,* 2015). Reintroduced populations may at times also be under very strong selection pressures due to the sudden change in environment they experience when founder individuals are released in new locations (e.g. Stockwell *et al.,* 2003; Reznick *et al.,* 2004). Hence, the selection pressures that were simulated are not biologically unrealistic and may occur in reintroduced populations.

Overall, for populations that experienced bottlenecks and reintroductions, the selection detection methods performed more poorly than has previously been shown for other demographic scenarios. Furthermore, the GEA approach had a higher performance than the *Fst*-like approaches. This is likely primarily due to the increased number of false positives for the *Fst*-like approach because of the difficulty in distinguishing drift-driven SNPs from those truly under selection. Importantly, the results confirm that the false positive rates quantified for non-bottlenecked populations cannot apply to studies on bottlenecked species. This has previous been assumed by studies on bottlenecked species (Pilot *et al.,* 2014; Funk *et al.,* 2016) and

may have lead to false confidence in loci identified as under selection. Based on these results, loci identified as under selection in bottlenecked populations should be viewed with caution, particularly those based on single selection detection methods. This is because the percentage of true positive loci may be as low as 1.8%.

Simulations and perfect knowledge of the environment

In this study, a consistently higher performance of GEA-based approaches was found and a true discovery rate of 78.2% was identified. In the analysis of the simulated data with a GEA approach, perfect knowledge of the environment was available because the true optimum was included in the analyses. Perfect knowledge of the true optimum is unlikely to arise in natural populations (De Mita *et al.,* 2013). As a result, the performance of GEA approaches in natural populations may be overestimated. Consequently, the effect of removing the true optimum from the GEA analysis on the number of true positives identified was quantified. Removing the true optimum from the analyses caused a minor reduction (14%) in the number of true positives in the candidate set identified by Bayenv2.0 and no reduction in those found by Baypass 2.1. These results are consistent with previous findings by De Mita *et al.* (2013), who identified only a marginal reduction in power when random noise was added to an environmental variable. The minor effect that was observed from removing the true optimum may be due to the presence of environmental variables that are correlated with the true optimum (Hoban *et al.,* 2016). These environmental variables may be sufficient to detect selection with a performance similar to those obtained in the analyses with the true optimum.

False positives and linkage in simulated data

The simulations included 30 chromosomes and neutral loci and loci under selection on were placed on placed these chromosomes. Consequently, it is possible that a subset of the false positive loci that were identified may have undergone allele frequency change due to linkage to a locus under selection. These can be considered either an error, because these loci themselves are not directly under selection, or they may be considered a marker for a region under selection, where the allele frequency change has arisen due to linkage to a locus under selection. As location information was not available for the simulated SNPs, the linkage disequilibrium between false positive loci and loci simulated under selection was estimated to identify the number

of false positive loci that are in high linkage disequilibrium with a true positive locus. Up to 3.9% of false positives loci were in moderate linkage disequilibrium (r^2>0.25) with a locus subject to selection. This percentage only contributes marginally to the high false discovery rate observed (up to 0.98). Thus, drift driven false positives are likely the primary force driving the high false discovery rate observed, and removing linked false positives does not substantially change the false positive rate.

Screen for selection with Alpine ibex RADseq data

In the Alpine ibex RADseq dataset the triple positive *Fst*-like approach identified five outliers and the triple positive GEA approach identified zero loci as putatively under selection. Due to the considerable false discovery rate of the *Fst*-like triple positive approach (98.3%), it is likely that none of these five candidate SNPs is truly subject to selection. At the same time, given the high false negative rate of the triple positive GEA approach (Figure 3), the absence of SNPs identified by this approach does not imply that no loci are under selection. However, these high rates of false positives prevent us from identifying loci under selection with confidence in the Alpine ibex.

The double positive GEA approaches identified up to 14 loci putatively under selection. However, based on the simulations, only 29-50% of loci identified by this approach are likely to be true positives. Therefore, these putatively selected loci should be viewed with caution. Interestingly, the significant environmental correlations observed with the loci putatively under selection in the Alpine ibex are most frequently related to winter conditions. Winter conditions may exert substantial selection pressure on Alpine ibex because winter climate has previously been shown to drive population fluctuations in Swiss Alpine ibex populations (Grøtan *et al.,* 2008) and in the Gran Paradiso region (Jacobsen *et al.,* 2004). However, at present there is only weak evidence for selection at the genetic level.

Selection detection after a bottleneck: A summary

Recent studies have begun to apply selection detection methods to identify signatures of selection in post-bottlenecked populations (Pilot *et al.,* 2014; Funk *et al.,* 2016; Brüniche-Olsen *et al*., 2016). While it is widely acknowledged that such studies may suffer high false positive rates (Foll and Gaggiotti, 2008; Brüniche-Olsen *et al*., 2016), the true discovery rate, the false discovery rate, the rate of false negatives and the actual rate of false positives are often unknown (though see Brüniche-Olsen *et al*.,

2016). In this study, it was demonstrated that only 1.3% to 13.1% of SNPs identified by any single selection detection approach may be under selection if the true discovery rate for loci of the largest allelic value is considered. Furthermore the false discovery rate can exceed 98% and the false positive rate can reach 1.76% when using the default conditions of a single selection detection approach in bottlenecked populations. Despite this, a true discovery rate up to 78.2% can also be reached if a GEA approach and multiple selection detection methods are used. The results emphasize that the details of demographic histories are important determinants of the success of selection detection methods in a particular study system. Simulations are helpful in determining the power of selection detection if the demography is known. Alternatively if time series samples are available it may sometimes be possible to disentangle selection and drift by comparing the results of single time point selection detection methods like those used here, to allele frequency changes over time (Brüniche-Olsen *et al*., 2016). In conclusion, though further attempts to identify selection after a bottleneck should be undertaken with care, they should neither be dismissed nor bottlenecked populations completely removed due to a high risk of false positive results.

References

Ardlie, K. G., Kruglyak, L., & Seielstad, M. (2002). Patterns of Linkage Disequilibrium in the Human Genome. Nature Reviews Genetics, 3(4), 299–309. doi.org/10.1038/nrg777

BAFU (2016) Ibex Colonies. Retrieved from: ch.bafu.fauna-steinbockkolonien

BAFU (2015) Abschuss Steinbock, ganze Schweiz: 2000-2015. Retrieved from: www.wild.uzh.ch/jagdst/index.php

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A.,…Johnson, E.A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3:e3376

Bamshad, M., & Wooding, S. P. (2003). Signatures of natural selection in the human genome. Nature Reviews Genetics, 4(2), 99–111. https://doi.org/10.1038/nrg999

Beaumont, M.A., & Balding, D. J. (2004). Identifying adaptive genetic divergence among populations from genome scans. Molecular Ecology, 13(4), 969–980. https://doi.org/10.1111/j.1365-294X.2004.02125.x

Beaumont, M. A., & Nichols, R. A. (1996). Evaluating loci for use in the genetic analysis of population structure. Proceedings of the Royal Society of London, Series B, 263(1377), 1619–1626. https://doi.org/10.1098/rspb.1996.0237

Berg, J. J., & Coop, G. (2014). A Population Genetic Signal of Polygenic Adaptation. PLoS Genetics, 10(8).1-25. https://doi.org/10.1371/journal.pgen.1004412

Berry, A., & Kreitman, M., (1993) Molecular analysis of an allozyme cline: alcohol dehydrogenase in Drosophila melanogaster on the east coast of North America. Genetics, 134, 869–893

Biebach, I., & Keller, L.F. (2009). A strong genetic footprint of the re-introduction history of Alpine ibex (*Capra ibex ibex*). Molecular Ecology. 18, 5046–5058. 10.1111/j.1365-294X.2009.04420.x

Biswas, S., & Akey, J. M. (2006). Genomic insights into positive selection. Trends in Genetics, 22(8), 437–446. https://doi.org/10.1016/j.tig.2006.06.005

Bonhomme, M., Chevalet, C., Servin, B., Boitard, S., Abdallah, J., Blott, S., SanCristobal, M. (2010). Detecting selection in population trees: The Lewontin and Krakauer test extended. Genetics, 186(1), 241–262. https://doi.org/10.1534/genetics.110.117275

Brüniche-Olsen, A., Austin, J. J., Jones, M. E., Holland, B. R., & Burridge, C. P. (2016). Detecting selection on temporal and spatial scales: A genomic time-series assessment of selective responses to devil facial tumor disease. PLoS ONE, 11(3), 1–15. doi.org/10.1371/journal.pone.0147875

Butchart, S.H.M., Walpole, M., Collen B., van Strien A., Scharlemnn J.P.W., Almond E.A, ... Watson, R. (2010). Global Biodiversity: Indicators of Recent Declines. Science, 328, 1164–1169. https://doi.org/10.1126/science.1187512

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., … Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly, 6, 80–92. 10.4161/fly.19695

Coop, G., Witonsky, D., Rienzo, A. D., & Pritchard, J. K. (2010). Using Environmental Correlations to Identify Loci Underlying Local Adaptation. Genetics, 185(4), 1411-1423 doi.org/10.1534/genetics.110.114819

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.,A., …Wang, J. (2011). The variant call format and VCFtools. Bioinformatics, 27,2156–8. 10.1093/bioinformatics/btr330

De Mita, S., Thuillet, A. C., Gay, L., Ahmadi, N., Manel, S., Ronfort, J., & Vigouroux, Y. (2013). Detecting selection along environmental gradients: Analysis of eight methods and their effectiveness for outbreeding and selfing populations. Molecular Ecology, 22(5), 1383–1399. https://doi.org/10.1111/mec.12182

Delcher, A.L., Phillippy, A., Carlton, J., & Salzberg, S.L. (2002). Fast algorithms for large-scale genome alignment and comparison. Nucleic Acids Research, 30, 2478–2483. 10.1093/nar/30.11.2478

Dirzo, R., Young, H. S., Galetti, M., Ceballos, G., Isaac, N. J. B., & Collen, B. (2014). Defaunation in the Anthropocene. Science, 345(6195), 401–406. https://doi.org/10.1126/science.1251817

Dong, Y., Xie, M., Jiang, Y., Xiao, N., Du, X., Zhang, W., ... Wang, W. (2013). Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). Nature Biotechnology, 31,135–41. doi:10.1038/nbt.2478

Duforet-Frebourg, N., Bazin, E., & Blum, M. G. B. (2014). Genome scans for detecting footprints of local adaptation using a Bayesian factor model. Molecular Biology and Evolution, 31(9), 2483–2495. https://doi.org/10.1093/molbev/msu182

Epstein, B., Jones, M., Hamede, R., Hendricks, S., McCallum, H., Murchison, E. P., … Storfer, A. (2016). Rapid evolutionary response to a transmissible cancer in Tasmanian devils. Nature Communications, 7, 12684. https://doi.org/10.1038/ncomms12684

Etter, P.D., Bassham, S., Hohenlohe, P.A., Johnson, E.A., & Cresko, W.A. (2011). Molecular Methods for Evolutionary Genetics. In: Orgogozo V, Rockman MV, editors. Vol. 772. Totowa, NJ: Humana Press. p. 157–178.

Excoffier, L., Hofer, T., & Foll, M. (2009). Detecting loci under selection in a hierarchically structured population. Heredity, 103(4), 285–98. https://doi.org/10.1038/hdy.2009.74

Fay, J. C., & Wu, C. I. (2000). Hitchhiking under positive Darwinian selection. Genetics, 155(3), 1405–1413.

Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. Genetics, 180(2), 977–993. https://doi.org/10.1534/genetics.108.092221

Frankham, R., Ballou, J., & Brisco, D. (2010). Introduction to Conservation Genetics (2nd ed.). UK: Cambridge University Press

Frichot, E., Schoville, S. D., Bouchard, G., François, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. Molecular Biology and Evolution, 30(7), 1687–1699. https://doi.org/10.1093/molbev/mst063

Funk, W. C., Lovich, R. E., Hohenlohe, P. A., Hofman, C. A., Morrison, S. A., Sillett, T. S., … Andelt, W. F. (2016). Adaptive divergence despite strong genetic drift: genomic analysis of the evolutionary mechanisms causing genetic differentiation in the island fox (*Urocyon littoralis*). Molecular Ecology, 25(10), 2176–2194. https://doi.org/10.1111/mec.13605

Gautier, M. (2015a). Genome-Wide Scan for Adaptive Divergence and Association with Population-Specific Covariates. Genetics 201,1555–79. 10.1534/genetics.115.181453

Gautier, M. (2015b). Baypass version 2.1 User Manual. Retrieved from: http://www1.montpellier.inra.fr/CBGP/software/baypass/index.html

Grøtan, V., Saether, B.E., Filli, F., & Engen, S. (2007). Effects of climate on population fluctuations of ibex. Global Change Biology, 14, 218–228 10.1111/j.1365-2486.2007.01484

Guillaume, F., & Rougemont, J. (2006). Nemo: an evolutionary and population genetics programming framework. Bioinformatics, 22, 2556–2557 10.1093/bioinformatics/btl415

Günther, T., & Coop, G. (2013). Robust identification of local adaptation from allele frequencies. Genetics, 195(1), 205–220. https://doi.org/10.1534/genetics.113.152462

Haddrill, P. R., Thornton, K. R., Charlesworth, B., & Andolfatto, P. (2005). Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. Genome Research, 15(6), 790–799. https://doi.org/10.1101/gr.3541005

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., … Whitlock, M. C. (2016). Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. The American Naturalist, 188(4), 379–397. https://doi.org/10.1086/688018

Hoeck, P. E. A., Bollmer, J. L., Parker, P. G., & Keller, L. F. (2010). Differentiation with drift: a spatio-temporal genetic analysis of Galapagos mockingbird populations (*Mimus spp*.). Philosophical Transactions of the Royal Society B: Biological Sciences, 365(1543), 1127–1138. https://doi.org/10.1098/rstb.2009.0311

Hofer, T., Ray, N., Wegmann, D., & Excoffier, L. (2009). Large allele frequency differences between human continental groups are more likely to have occurred by drift during range expansions than by selection. Annals of Human Genetics, 73(1), 95–108. https://doi.org/10.1111/j.1469-1809.2008.00489.x

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genetics, 6(2). https://doi.org/10.1371/journal.pgen.1000862

Jacobson, A., & Provenzale, A. (2004). Climate forcing and density dependence in a mountain ungulate population. Ecology. 85,1598–1610. doi/10.1890/02-0753

Kahle, D., & Wickham, H. (2013) ggmap: Spatial Visualization with ggplot2. The R Journal, 5(1), 144-161.

Karlsson, S., & Moen, T. (2010). The power to detect artificial selection acting on single loci in recently domesticated species. BMC Research Notes, 3, 232. https://doi.org/10.1186/1756-0500-3-232

Keller, S. R., Sowell, D. R., Neiman, M., Wolfe, L. M., & Taylor, D. R. (2009). Adaptation and colonization history affect the evolution of clines in two introduced species. New Phytologist, 183(3), 678–690. https://doi.org/10.1111/j.1469-8137.2009.02892.x

Kemper, K. E., Saxton, S. J., Bolormaa, S., Hayes, B. J., & Goddard, M. E. (2014). Selection for complex traits leaves little or no classic signatures of selection. BMC Genomics, 15(1), 246. https://doi.org/10.1186/1471-2164-15-246

Kimura, M. (1955a). Solution of a Process of Random Genetic Drift With a Continuous Model. Proceedings of the National Academy of Sciences of the United States of America, 41(3), 144–50. https://doi.org/10.1073/pnas.41.3.144

Kimura, M. (1955b). Random Genetic Drift in Multi-Allelic Locus. Evolution, 9(4), 419–435. 10.2307/2405476

Klopfstein, S., Currat, M., & Excoffier, L. (2006). The fate of mutations surfing on the wave of a range expansion. Molecular Biology and Evolution, 23(3), 482–490. https://doi.org/10.1093/molbev/msj057

Lamichhaney, S., Han, F., & Berglund, J. (2016). A beak size locus in Darwin's finches facilitated character displacement during a drought. Science 352(6284):470-474. 10.1126/science.aad8786

Latter, B. D. (1973). The island model of population differentiation: a general solution. Genetics, 73(1), 147–157.

Lescak, E.A., Bassham, S.L., Catchen, J., Gelmond, O., Sherbick, M.L., von Hippel, F.A., & Cresko, W.A. (2015). Evolution of stickleback in 50 years on earthquake-uplifted islands. PNAS. 112:E7204–12. 10.1073/pnas.1512020112

Lewontin, R. C., & Krakauer, J. (1973). Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics, 74(1), 175–195.

Linnen, C. R., Poh, Y.P., Peterson, B. K., Barrett, R. D. H., Larson, J. G., Jensen, J. D., & Hoekstra, H. E. (2013). Adaptive Evolution of Multiple Traits Through Multiple Mutations at a Single Gene. Science, 339(6125), 1312–1316. https://doi.org/10.1126/science.1233213

Lischer, H.E.L., & Excoffier, L. (2012). PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics. 28, 298–299. 10.1093/bioinformatics/btr642

Lotterhos, K.E., & Whitlock, M.C. (2014). Evaluation of demographic history and neutral parameterization on the performance of Fst outlier tests. Molecular Ecology. 23, 2178–2192. 10.1111/mec.12725

Lotterhos, K. E., & Whitlock, M. C. (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. Molecular Ecology, 24(5), 1031–46. https://doi.org/10.1111/mec.13100

Lynch D. (1996) A quantitative-genetic perspective on conservation issues, pp. 471-501. In J. Avise and J. Hamrick (eds.) Conservation Genetics: Case Histories from Nature. Chapman and Hall, York

MeteoSwiss (2016) meteoswiss.admin.ch

Nadeau, S., Meirmans, P. G., Aitken, S. N., Ritland, K., & Isabel, N. (2016). The challenge of separating signatures of local adaptation from those of isolation by distance and colonization history: The case of two white pines. Ecology and Evolution, 6(24), 8649–8664. https://doi.org/10.1002/ece3.2550

Narum, S. R., & Hess, J. E. (2011). Comparison of F(ST) outlier tests for SNP loci under selection. Molecular Ecology Resources, 11 Suppl 1, 184–94. https://doi.org/10.1111/j.1755-0998.2011.02987.x

Nielsen, R., Hellmann, I., Hubisz, M., Bustamante, C., & Clark, A. G. (2007). Recent and ongoing selection in the human genome. Nature Reviews Genetics, 8(11), 857–868. https://doi.org/10.1038/nrg2187

Pilot, M., Greco, C., vonHoldt, B. M., Jędrzejewska, B., Randi, E., Jędrzejewski, W., … Wayne, R. K. (2014). Genome-wide signatures of population bottlenecks and diversifying selection in European wolves. Heredity, 112(4), 428–442. https://doi.org/10.1038/hdy.2013.122

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. Molecular Ecology, 24(17), 4348–4370. https://doi.org/10.1111/mec.13322

Reznick, D., Rodd, H., & Nunney, L. (2004) Empirical evidence for rapid evolution. In: Ferrière, R., Dieckmann, U., Couvet, D. ed. Evolutionary Conservation Biology. UK: Cambridge University Press

Robertson, A. (1975a). Gene frequency distributions as a test of selective neutrality. Genetics, 81(4), 775–785.

Robertson, A. (1975b). Letters to the editors: Remarks on the Lewontin-Krakauer test. Genetics, 80(2), 396.

Sabeti, P. C. (2006). Positive Natural Selection in the Human Lineage. Science, 312(5780), 1614–1620. https://doi.org/10.1126/science.1124309

Seddon, P.J., Strauss, W.M., & Innes, J. (2012). Animal Translocations: What Are They and Why Do We Do Them? In: Ewen, J.G., Armstrong, D.P., Parker, K.A., Seddon, P.J. ed. Reintroduction Biology. Chichester: Wiley. Pg.1-32

Stockwell, C.A., Hendry, A.P., & Kinnison, M.T. (2003). Contemporary evolution meets conservation biology. Trends in Ecology Evolution. 18, 94–101 dx.doi.org/10.1016/S0169-5347(02)00044-7

Storey, J. D., & Tibshirani, R. (2003). Statistical significance for genome-wide studies. Proceedings of the National Academy of Sciences, 100(16), 9440–9445. https://doi.org/10.1073/pnas.1530509100

Shackleton, D., & Group ISCS. (1997). Wild Sheep and Goats and their Relatives. Status Survey and Conservation Action Plan for *Caprinae*. IUCN, Gland, Switzerland and Cambridge.

Stuwe, M., & Grodinsky, C. (1987). Reproductive biology of captive Alpine ibex (*Capra i. ibex*). Zoo Biology. 6, 331–339. 10.1002/zoo.1430060407

Stuwe, M., & Nievergelt, B. (1991). Recovery of Alpine ibex from near extinction – the result of effective protection, captive breeding, and reintroductions. Applied Animal Behaviour Science. 29,379–387.

Thornton, K., & Andolfatto, P. (2006). Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Netherlands population of *Drosophila melanogaster*. Genetics, 172(3), 1607–1619. https://doi.org/10.1534/genetics.105.048223

Umina, P. A., Weeks, A. R., Kearney, M. R., McKechnie, S. W., & Hoffmann, A. A. (2005) A rapid shift in a classic clinal pattern in Drosophila reflecting climate change. Science, 308, 691–693

Vitalis, R., Dawson, K., Boursot, P., & Belkhir, K. (2003). Computer Note: DetSel 1.0: A Computer Program to Detect Markers Responding to Selection. Journal of Heredity, 94(5), 429–431. https://doi.org/10.1093/jhered/esg083

Whitlock, M. C., & Lotterhos, K. E. (2015a). Reliable Detection of Loci Responsible for Local Adaptation: Inference of a Null Model through Trimming the Distribution of *FST*. The American Naturalist, 186(S1), S24–S36. https://doi.org/10.1086/682949

Whitlock, M. C., & Lotterhos, K. E. (2015b). OutFLANK Finding *Fst* outliers with an inferred neutral distribution. https://github.com/whitlock/OutFLANK

Wickham, H. (2009) ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag

Wright, S. (1949). The Genetical Structure of Populations. Annals of Eugenics, 15(1), 323–354. https://doi.org/10.1111/j.1469-1809.1949.tb02451.x

Supplementary Material

Section S1:

Simulations of neutral loci and loci under selection:

Nemo (version 2.3.44; Guillaume and Rougemont, 2006) was used to create forward simulations where individuals follow a user-determined life cycle of reproduction, migration and breeding partner numbers. Genotypes were generated for each individual and were sampled at the end of the simulations. 26 populations were simulated, representing the 23 focal Swiss Alpine ibex populations, the Gran Paradiso (source population) and two additional captive populations used as founder sources. To reach mutation-drift equilibrium and enable a build up of linkage disequilibrium, an initial run of 50,000 generations was conducted for the Gran Paradiso population. These 50,000 generations were chosen based on exploratory analyses on the time necessary to reach equilibrium.

After reaching equilibrium, a bottleneck was applied to mirror the global decline in Alpine ibex. The Gran Paradiso population was reduced to 80 individuals and then allowed to regrow over 7 generations, after which reintroductions began. Nemo does not currently support overlapping generations, therefore the time from the first reintroduction in 1906 to 2015 was divided into discrete generations of eight years. Eight years corresponds to the generation time of Alpine ibex (Stuwe and Grodinsky, 1987). All individuals that were reintroduced in each eight-year period were represented as one migration event in the corresponding Nemo generation. The period from 1906 to 2015 corresponds to 13.6 generations. Thus to ensure that all translocations were included in the simulations, 14 generations of reintroductions were simulated, the last five-year period correspond to one full generation. In addition, a final generation was allowed after the 14 generations of reintroductions, so that the final reintroduced individuals could reproduce before sampling. No migration outside of recorded translocations has been detected genetically (Aeschbacher *et al.,* 2013). Therefore no additional migrations were allowed between simulated populations. Five minor contributing source populations in were also ignored in the simulations, as this would have created further complexity, arising from the need to simulate their founders in turn. The effect of ignoring these populations should be negligible, because genetically very similar populations were available to use as source population for the simulated translocations. Only very minor reintroductions (fewer than six individuals) from these non-simulated populations were ignored.

Migration events within Nemo are determined from migration rates. Specifically the user determines the proportion of the source population individuals that migrate. Consequently, for any simulated movement of individuals that represented a translocation, the proportion of individuals that migrated was monitored to ensure this was as similar as possible to the size of the translocation event. One source population was founded a generation early to ensure migration from that population could occur at the correct generation. Populations were permitted to grow to a pre-determined maximum size, after which individuals were randomly removed to keep the population size constant. The maximum size was the largest census size observed between the start of recorded population sizes and the year 2006. 2006 was chosen as it corresponded to the start of the final generation within simulations and the start of the genetic sampling. However, it should be noted that Alpine ibex populations often attained their maximum population size well before 2006 in the wild. The maximum Alpine ibex population size used excluded individuals less than one-year-old where possible, because Nemo only regulates adult numbers (Guillaume and Rougemont, 2006). No census data was available for three populations, consequently their maximum size was set to 319 individuals, which is the mean maximum population size across the known wild Swiss populations. All individuals were given a mean lifetime fecundity of 9 offspring to prevent early population extinction in the reintroduction phase of the simulations. Reproductive success followed a Poisson distribution and males were randomly chosen for each female. In the wild, Alpine ibex have a strong skew in reproductive successes to a handful of older males in a population (Willisch *et al.,* 2012). Unfortunately, a similar option was not available in Nemo, which is why randomly chosen males were used instead.

Ten replicate starting simulations ('burn-ins') were conducted to ensure that populations reached mutation-drift equilibrium. After this time no new mutational input was allowed. These populations were then used as starting points for 15 bottleneck replicate simulations. For each simulation the evolution of 60,000 SNPs on 30 chromosomes of equal size was simulated. 60,000 SNPs ensured that a similar number of SNPs was polymorphic at the end of the simulations as in the empirical RADseq data set. SNPs were randomly distributed across a genetic map with a resolution of $1x\ 10^{-5}$ centimorgans. This created a genome similar in size and structure to that of the goat (Guillaume and Rougemont, 2006; Bickhart *et al.,* 2017). All SNPs were initially monomorphic and variation arose at a mutation rate (μ) equal

to 1 x 10^{-6} per locus and generation. This rate was chosen to ensure that a sufficient number of polymorphisms survived each bottleneck (Ségurel *et al.*, 2014). To simulate selection, 30 additional SNPs were placed individually on each chromosome randomly, each of these contributed additively to a single trait. Specifically, 10 SNPs were considered loci of large effect, with allelic values of 0.1 and were thus subject to strong selection. Ten SNPs with effects of 0.01 and 10 SNPs with effects of 0.001 were also included, representing loci of moderate-to-weak effects that were under weaker selection. During the burn-in a local trait optimum of zero was used to maintain alleles at the 30 loci under selection, which have both positive and negative contributions on the trait value. To maintain polymorphisms at the loci under selection a higher mutation rate of 1 x 10^{-4} was used. All loci, neutral and selected, had an equal likelihood of mutation under their respective mutation rates and were confined to just two alleles. Population optimum trait values for reintroduced populations were set to reflect the patterns seen in natural environmental conditions. Specifically, the direction of the optimum was based on the snow depth values of the real populations to create a realistic pattern of environmental variation. Simulated optima are shown in Table S1. Values were set close to the maximum or minimum trait values that can be achieved by the additively loci and therefore were either -2, 2 or 0.

For selection detection analyses a maximum of ten individuals from 23 populations were selected. Fixed SNPs were excluded from this analysis. Additionally, the three populations that were not sampled genetically but were simulated as founder sources were also excluded from the search for signals of positive selection in the simulated data. All genotypes were taken from the final generation of the simulations.

Table S1: *The true optima of the populations included in the simulations. Positive and negative values were based on the relative difference in snow depth between each population and the Gran Paradiso population. An optimum of zero was used for the burn-in.*

Population	Simulated optimum
Gran Paradiso	0
Wilderness Park Peter and Paul	-2
Wilderness Park Interlaken Harder	-2
Graue Hoerner	-2
Albris	0
Brienzer-Rothorn	0
Schwarmoench	-2
Wetterhorn	0
Mont Pleureur	-2
Justistal	0
Gross Lohner	-2
Alpstein	2
Val Bever	-2
Crap da Flem	-2
Flueela	0
Wittenberg	0
Arolla	-2
Bire-Oeschinen	-2
Creux du Van	0
Pilatus	0
Fluebrig	-2
Weisshorn	-2
Oberbauenstock	-2
Falknis	0
Tanay	-2
Churfirsten	2

Table S2: *Weather stations used for each population. Data included in the analyses are an average of each season over the years since each population's founding or when records began. Stations were chosen based on proximity and similarity in conditions.*

Colony	First Translocation	Weather Time Series Start	Station	Station Elevation (m)
Bire-Oeschinen	1961	1962	ABO	1320
Gross Lohner	1952	1953	ABO	1320
Schwarmoench	1924	1931	ABO	1320
Tanay	1977	1978	AIG	381
Albris	1920	1931	BEH	2307
Crap da Flem	1958	1959	ELM	965
Fluebrig	1962	1963	ELM	965
Graue Hoerner	1911	1912	ELM	965
Oberbauenstock	1969	1970	ENG	1037
Arolla	1960	1978	EVO	1825
Creux du Van	1961	1978	FRE	1205
Wetterhorn	1926	1931	GRH	1980
Mont Pleureur	1928	1928	GSB	2472
Wittenberg	1958	1978	MLS	1972
Brienzer-Rothorn	1921	1978	PIL	2106
Justistal	1949	1978	PIL	2106
Pilatus	1961	1978	PIL	2106
Alpstein	1955	1956	SAE	2502
Churfirsten	1984	1985	SAE	2502
Falknis	1970	1971	WFJ	2690
Flueela	1958	1959	WFJ	2690
Val Bever	1957	1978	SAM	1705
Weisshorn	1962	1962	ZER	1638

Section S2:
Weather Data:

To include a measure of the environmental conditions a population experienced, the environmental conditions of the closest meteorological weather station were averaged across years. If weather records began before a population was founded, data were averaged from the year after the first translocation event until October 2015. For populations where weather records began after the first translocation event, all data were used. Each environmental variable included was divided into three seasons: winter (November-April), summer (May-October), and late winter (February to April) (Buzzuto*, pers.com*). Late winter was included as it captures the period with highest snow depths. The environmental variables included in the analyses were: daily mean, maximum and minimum air temperature 2 meters above ground (°C), daily precipitation (mm), daily new snow accumulation (cm), and total snow height (cm) at 5:40 am (MeteoSwiss, 2016). In addition, the number of days a population experienced above the average snow depth of all populations, and above twice the average snow depth of all populations, were included. Mean values were calculated separately for each season and used in Bayenv 2.0 and Baypass2.1. For the simulated data sets the mean value of each environmental variable and season corresponding to the simulated population were used in the GEA approaches, as well as the simulated population's true optima, i.e. the environmental values to which the loci subject to selection were adapting.

Section S3:
RADseq bioinformatics steps:

DNA was extracted using the QIAGEN DNeasy Blood Tissue Kit (QIAGEN). DNA quantity was assessed using picogreen (QuantIT)(Ahn *et al.,* 1996) and a 1% agarose gel was used to check for DNA degradation. RAD libraries were constructed at the Genetic Diversity Centre, ETH Zurich following the Etter protocol (Etter *et al.,* 2011) with a SBF1 digest. Minor modifications were applied to the Etter protocol and are detailed in Grossen *et al.,* (2014). Samples were dual indexed and paired-end sequenced (number of samples per sequencing read length: 100bp=85, 125bp=30, 140bp=189) on an Illumina HiSeq at the Functional Genomics Centre of the University of Zurich (2012; n=85) and the Genomics Facility Basel, ETH Zurich (2015; n=219). To avoid confounding effects from lane success bias and library preparation biases, samples from individual populations were split across lanes and libraries where possible.

Sequence data were demultiplexed using the FASTX-toolkit (version 0.0.14; Person *et al*., 1997) with one partial overlap and two mismatches in the barcode sequence allowed, including deletions. Due to a single base insertion found before the inline barcode, unassigned reads were re-demultiplexed with one mismatch and partial overlap after trimming the first base pair. Demultiplexed reads were then cleaned with Trimmomatic (version 0.32: Bolger *et al.,* 2014) using the palindrome mode, which is suitable for the longer read lengths generated here and removes sequences of adaptors that may be present at the end of reads. A 16 base pair seed with two mismatches was allowed to find any adaptor sequences (Bolger *et al.,* 2014). Bases at the start and end of reads with a PHRED score less than three were also removed. In addition, reads were trimmed if a sliding window of four bases with a PHRED less than 15 arose. Reads less than 36 base pairs long were also discarded. All remaining reads were then mapped against the goat genome (01.Genome, scaffold file; Dong *et al.,* 2013) with Bowtie2 using the *sensitive* mode and an expected insert size of 50-800 base pairs between reads (version 2.2.5: Langmead and Salzberg, 2012). PCR duplicates were removed with Markduplicates in Picardtools (Broad, 2016). Variant nucleotides were called using Freebayes (version v0.9.21-19-gc003c1e: Garrison and Marth, 2012), evaluating a maximum of three alleles at each site to reduce computational time and without prior population assignment. Fixed differences between Alpine ibex and the goat genome were removed before filtering

using custom scripts. SNPs were then filtered with vcfilter (available in Freebayes) and vcftools 0.1.14 (Garrison, 2016; Danecek *et al*., 2011). SNPs were removed if the site quality to depth ratio was less than 0.25, if site quality was less than 1, or if mean mapping quality was less than 30. As recommended for Freebayes, sites with a quality score divided by the number of alternative allele observations of less than 10 were removed to discard poor quality sites. Similarly, SNPs with a genotyping quality of below 20, a phred score less than 40, and a depth of less than 8 reads were also excluded. SNP sites were then included only if the allelic balance of the SNP was between 0.25 and 0.75 or less than 0.01. The allelic balance is the ratio of reads that show each allele in heterozygote individuals. This filter ensured a balance ratio of reads showing each allele was present at a variant sites, and removed false SNP calls. SNPs with a mapping quality ratio of the two alleles less than 0.9 or greater than 1.05 were also removed to avoid strand bias, as recommended by dDocent (Puritz *et al.,* 2014). To avoid biases arising from different library lengths at the read-ends, SNPs were also removed if the alternative allele was not seen on more than one forward or reverse strand and if less than one read was present at either side of the SNP. Additionally, SNPs with more than two alleles were removed, because the selection detection programs could not accept non-biallelic loci. Finally, SNPs with a mean maximum read depth across all individuals greater than twice the mean depth were removed (Li, 2014) and those with a heterozygosity of greater than 0.6 were excluded to ensure that paralogs and duplicate regions were not present. The value of 0.6 was chosen, because under Hardy-Weinberg equilibrium heterozygosity above 0.5 would not be expected, and may signal errors in the data.

Section S4:
Performance for *Fst*-like and GEA approaches:

Table S3: *The composition of loci identified as being under selection by different detection methods i.e. the true and false discovery rates. Triple positives are loci identified by three selection detection methods, double positives are identified by two. Fst-like approaches detect selection using high allelic differentiation and GEA approaches detect selection through identifying repeated associations between allele frequencies with an environmental variable. Columns labelled 0.1, 0.01,0.001 indicate true positives for loci of the corresponding allelic value. False positive indicates neutral loci incorrectly identified here as under selection. Shown is the percentage of all loci identified as under selection for each type of locus as an average across the 11 simulated data sets (± one standard deviation). Those loci with an allelic value of 0.1 are under a selection pressure of 0.13. Those with a value of 0.01 or 0.001 are under a selection pressure of 0.013 or 0.0013.*

	Percentage of Each Locus Type Identified as Under Selection (%)			
Test	**0.1**	**0.01**	**0.001**	**False Positive**
Triple positive Fst-like	1.7±3.7	0±0	0±0	98.3±3.7
Triple positive GEA	78.2±33.7	0±0	0±0	21.8±33.7
Double positive Bayenv Baypass Fst-like	1.7±3.7	0±0	0±0	98.3±3.7
Double positive Baypass OutFLANK Fst-like	9.1±12.7	0±0	0±0	90.9±12.7
Double positive Bayenv OutFLANK Fst-like	2±1.9	0±0	0.3±0.6	97.8±2.2
Double positive Bayenv Baypass GEA	49.7±25.2	0±0	0±0	50.3±25.2
Double positive Baypass OutFLANK GEA	35.7±17	0±0	0±0	64.3±17
Double positive Bayenv OutFLANK GEA	28.7±15	0±0	0±0	71.3±15
Bayenv Fst-like	1.3±1.3	0.2±0.5	0.3±0.5	98.2±1.9
Baypass Fst-like	9.3±12.5	0±0	0±0	90.7±12.5
OutFLANK right tailed Fst-like	1.7±1.1	0.1±0.2	0.1±0.2	98.1±1.3
Bayenv GEA	13.1±4.4	0.1±0.3	0±0	86.8±4.3
Baypass GEA	7.2±2.9	0±0	0.3±1	92.5±3.4

Table S4: *The true positive rate for each type of selected locus. Loci that were simulated to be under selection are divided into three classes depending on their allelic value, the contribution alleles at the loci can make to the trait under selection. Those with a value of 0.1 are under a selection pressure of 0.31. Those with a value of 0.01 or 0.001 are under a selection pressure of 0.013 or 0.0013. Ten loci of each class were simulated. The true positive rate is the number of loci correctly identified as under selection divided by the total number of loci of each type simulated (10). Data correspond to averages across the 11 simulations (± one standard deviation). The values shown subtracted from 100 give the false negative rates.*

	True Positive Rate Per Locus Type (%)		
Method	**0.1**	**0.01**	**0.001**
Triple positive Fst-like	1.8±4	0±0	0±0
Triple positive GEA	16.4±9.2	0±0	0±0
Double positive Bayenv Baypass Fst-like	1.8±4	0±0	0±0
Double positive Baypass OutFLANK Fst-like	7.3±7.9	0±0	0±0
Double positive Bayenv OutFLANK Fst-like	10.9±10.4	0±0	1.8±4.1
Double positive Bayenv Baypass GEA	23.6±9.2	0±0	0±0
Double positive Baypass OutFLANK GEA	17.3±11	0±0	0±0
Double positive Bayenv OutFLANK GEA	30.0±22.8	0±0	0±0
Bayenv Fst-like	10.9±10.4	1.8±4.1	2.7±4.7
Baypass Fst-like	8.2±7.5	0±0	0±0
OutFLANK right tailed Fst-like	35.5±24.6	0.9±3	1.8±04.1
Bayenv GEA	60±14.1	0.9±3	0±0
Baypass GEA	26.4±11.2	0±0	0.9±3

Table S5: *Mean false positive rate for each selection detection method. False positives are neutral loci incorrectly identified as being under selection. The false positive rate is the number of false positives identified divided by the total number of neutral loci. Data are the averages across 11 simulations (± one standard deviation).*

Method	**Mean False Positive Rate (%)**
Triple positive Fst-like	0.13±0.09
Triple positive GEA	0.01±0.01
Double positive Bayenv Baypass Fst-like	0.14±0.09
Double positive Baypass OutFLANK Fst-like	0.23±0.13
Double positive Bayenv OutFLANK Fst-like	1.15±0.2
Double positive Bayenv Baypass GEA	0.05±0.03
Double positive Baypass OutFLANK GEA	0.06±0.03
Double positive Bayenv OutFLANK GEA	0.14±0.12
Baypass Fst-like	0.25±0.16
Bayenv Fst-like	1.76±0.12
OutFLANK right tailed Fst-like	4.13±0.59
Baypass GEA	0.75±0.25
Bayenv GEA	0.91±0.39

Section S5:
Linkage of false positives to selected loci:

Table S6: *The number of false positive neutral loci in high linkage disequilibrium (r^2>0.25) with a simulated selected locus. These loci are initially counted as false positives but show a true signature of selection due to linkage with a selected region. Values shown in the second and third column are the average across 11 simulations (± one standard deviation).*

Method Applied	Number of False Positive loci	Number of loci with LD values above 0.25 with selected loci	Percentage of false positives with high LD (%)
Triple positive Fst-like	6.5±4.4	0.2±0.4	2.8
Triple positive GEA	0.5±0.7	0±0	0
Double positive Bayenv Baypass Fst-like	6.5±4.5	0.2±0.4	2.7
Double positive Baypass OutFLANK Fst-like	11±6.5	0.4±0.7	3.3
Double positive Bayenv OutFLANK Fst-like	55.6±9.8	1.8±2.3	3.3
Double positive Bayenv Baypass GEA	2.5±1.6	0±0	0
Double positive Baypass OutFLANK GEA	3.1±1.3	0±0	0
Double positive Bayenv OutFLANK GEA	6.9±6.1	0.3±0.5	3.9
Baypass Fst-like	12.1±7.6	0.4±0.7	3
Bayenv Fst-like	84.6±4.6	1.9±2.3	2.3
OutFLANK right tailed p-value	199.4±30	4.6±3.4	2.3
Baypass GEA	36.3±11.5	0.2±0.4	0.5
Bayenv GEA	43.7±18.8	0.5±0.5	1.3

Section S6:
Double positive environmentally correlated SNPs

Table S7: *Outlier SNPs identified by the double positive environmental correlation tests for the Alpine ibex RADseq data. Listed are the significantly associated environmental variables and the SNP location as annotated by SNPeff.*

Chromosome	Position (bp)	Method	SNP location	Environmental Correlation
Chr 6	86864948	OutFLANK Baypass	intergenic	Winter daily precipitation Winter total snow depth
Chr 7	93296232	OutFLANK Baypass	intergenic	Daily mean temperature winter short
Chr 11	13183701	OutFLANK Baypass	intergenic	Daily maximum temperature in winter
Chr 14	2958961	OutFLANK Baypass	intergenic	Daily mean temperature winter short
Chr 16	70559366	OutFLANK Baypass	intron	Winter daily precipitation Winter mean temperature
Chr 17	4592929	OutFLANK Baypass	intergenic	Maximum temperature in winter
Chr 2	48478378	OutFLANK Baypass	intergenic	Daily mean temperature winter short
Chr 20	57945948	OutFLANK Baypass	intron	Daily mean temperature winter short
Chr 13	68115192	OutFLANK Bayenv	downstream	Days above twice the mean snow depth in summer and winter Days above the mean snow depth in winter summer and winter short Minimum air temperature in summer, winter and winter short Mean air temperature in summer, winter and winter short Maximum air temperature in winter and winter short Total snow in winter and winter short New snow depth in summer
Chr 29	32374375	Bayenv Baypass	intergenic	Winter mean temperature, Winter minimum temperature, Days above twice the mean snow depth
Chr 7	41543007	Bayenv Baypass	intergenic	Snow height in winter, late winter, summer
Chr 15	48848628	Bayenv Baypass	intron	Snow height in summer
Chr 15	75395333	Bayenv Baypass	downstream	Height of fresh snow in winter and summer, days above twice the mean snow depth
Chr 15	78880490	Bayenv Baypass	intron	Days above the mean snow depth

Section S7:
Triple positive Fst-like outlier SNPs.

Table S8: *The Fst derived outliers found in the RADseq data set using the triple positive method.*

Chromosome	Position (bp)	Method	Region
Chr 1	79503962	Triple positive Fst-derived	intergenic
Chr 2	36378570	Triple positive Fst-derived	intron
Chr 7	9372371	Triple positive Fst-derived	intron
Chr 16	2983926	Triple positive Fst-derived	intron
Chr 24	46120599	Triple positive Fst-derived	intergenic

Supplementary Material References

Aeschbacher, S., Futschik, A., & Beaumont, M.A. (2013). Approximate Bayesian computation for modular inference problems with many parameters: the example of migration rates. Molecular ecology, 22(4), 987–1002. doi:10.1111/mec.12165

Ahn, S.J., Costa, J., & Emanuel, J.R. (1996). PicoGreen quantitation of DNA: effective evaluation of samples pre- or post-PCR. Nucleic Acids Res. 24:2623–2625.

Bickhart, D.M., Rosen, B.J., Koren, S., Sayre, B.L., Hastie, A.R.,…Smith, T.P.L. (2016) Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. Nature Genetics. 49:643-650 doi:10.1038/ng.3802

Bolger, A.M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30:2114–2120. 10.1093/bioinformatics/btu170

Broad (2016) Picard. Retrieved from http://broadinstitute.github.io/picard.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., …Wang, J. (2011). The variant call format and VCFtools. Bioinformatics 27:2156–8. 10.1093/bioinformatics/btr330

Dong, Y., Xie, M., Jiang, Y., Xiao, N., Du, X., Zhang, W., ... Wang, W. (2013). Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (Capra hircus). Nature Biotechnology. 31:135–41. doi:10.1038/nbt.2478

Etter, P.D., Bassham, S., Hohenlohe, P.A., Johnson, E.A., & Cresko, W.A. (2011). Molecular Methods for Evolutionary Genetics. In: Orgogozo V, Rockman MV, editors. Vol. 772. Totowa, NJ: Humana Press. p. 157–178.

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. arXiv:1–9

Garrison, E. (2016) vcflib: A C++ library for parsing and manipulating VCF files. Retrieved from: https://github.com/vcflib/vcflib

Grossen, C., Keller, L., Biebach, I., & Croll, D. (2014). Introgression from domestic goat generated variation at the major histocompatibility complex of Alpine ibex. PLoS Geneteics. 10(6),1-19, 10:e1004438

Guillaume, F., & Rougemont, J. (2006). Nemo: an evolutionary and population genetics programming framework. Bioinformatics. 22:2556–7 10.1093/bioinformatics/btl415

Langmead, B., & Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods 9:357–9.

Li, H. 2014. Toward better understanding of artefacts in variant calling from high-coverage samples. Bioinformatics 30:2843–51. 10.1093/bioinformatics/btu356

MeteoSwiss (2016) meteoswiss.admin.ch

Pearson, W.R., Wood, T., Zhang, Z., & Miller, W. (1997). Comparison of DNA Sequences with Protein Sequences. Genomics. 36:24–36. dx.doi.org/10.1006/geno.1997.4995

Puritz, J.B., Hollenbeck, C.M., & Gold, J.R. (2014). dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. PeerJ 2:e431. 10.7717/peerj.431

Ségurel, L., Wyman, M. J., & Przeworski, M. (2014). Determinants of Mutation Rate Variation in the Human Germline. Annual Review of Genomics and Human Genetics, 15(1), 47–70. https://doi.org/10.1146/annurev-genom-031714-125740

Stuwe, M., & Grodinsky, C. (1987). Reproductive biology of captive Alpine ibex (*Capra i. ibex*). Zoo Biology. 6,331–339. 10.1002/zoo.1430060407

Willisch, C., I. Biebach, U. Koller, T. Bucher, N. Marreros,…Neuhaus, P. (2012) Male reproductive pattern in a polygynous ungulate with a slow life-history: the role of age, social status and alternative mating tactics. Evol. Ecol. 26,187–206